Received: 8 June 2018 | Revised: 15 January 2019 | Accepted: 15 January 2019

DOI: 10.1002/asmb.2434

RESEARCH ARTICLE

WILEY

Bayesian semiparametric Markov switching stochastic volatility model

Audronė Virbickaitė[1] | Hedibert F. Lopes[2]

[1]Department of Applied Economics, Universitat de les Illes Balears (UIB), Palma de Mallorca, Spain

[2]Insper Institute of Education and Research, Sao Paulo, Brazil

Correspondence
Audronė Virbickaitė, Department of Applied Economics, Universitat de les Illes Balears (UIB), 07122 Palma de Mallorca, Spain.
Email: audrone.virbickaite@uib.es

Abstract

This paper proposes a novel Bayesian semiparametric stochastic volatility model with Markov switching regimes for modeling the dynamics of the financial returns. The distribution of the error term of the returns is modeled as an infinite mixture of Normals; meanwhile, the intercept of the volatility equation is allowed to switch between two regimes. The proposed model is estimated using a novel sequential Monte Carlo method called particle learning that is especially well suited for state-space models. The model is tested on simulated data and, using real financial times series, compared to a model without the Markov switching regimes. The results show that including a Markov switching specification provides higher predictive power for the entire distribution, as well as in the tails of the distribution. Finally, the estimate of the persistence parameter decreases significantly, a finding consistent with previous empirical studies.

KEYWORDS

Bayes factor, Dirichlet process mixture, particle learning, sequential Monte Carlo

1 | INTRODUCTION

Volatility modeling has been of great interest in the last decades, especially after the recent financial crisis when the standard models failed to explain and predict the events that occurred in the financial markets. The two benchmark approaches to model volatility are based on the autoregressive conditional heteroscedasticity (ARCH)–type models, proposed by Engle,[1] and the stochastic volatility (SV) type models, proposed by Taylor.[2] The models differ in the underlying assumptions of the observability of the volatility. In ARCH-type models, the volatility is deterministic and observable, whereas, in the SV type models, the volatility states are latent and stochastic. By allowing for the volatility states to be stochastic, SV models provide more flexibility than the generalized ARCH (GARCH[3]) specifications (see the work of Broto and Ruiz[4] for example).

The SV model, as introduced by Taylor,[2] assumes the distribution of the error term of the returns to be Normal. Normal distribution was also considered by other works,[5-8] just to name a few. However, many empirical studies have shown that the returns exhibit heavy-tailed behavior (see related works[9-11] for example). One possibility, instead of Normal distribution, is to employ a distribution that allows for fat tails. The student-t distribution was used by other works[10-15]; the Normal-inverse Gaussian by Barndorff-Nielsen[16]; the mixture of Normals by Mahieu and Schotman[17]; and the generalized error distribution by Liesenfeld and Richard,[18] among many others. Another possibility is to abandon parametric assumptions for the distribution of the returns altogether and consider a semiparametric SV model.* In

*In some papers, such models are referred to as nonparametric SV models

such model, the volatility equation maintains the parametric form, whereas the distribution of the returns is modeled nonparametrically.

The Bayesian semiparametric SV models have become rather popular in the last decade (see other works[19,20] and[21,22] for univariate SV models). Zaharieva et al[23] proposed a multivariate extension with the nonparametric errors. In these works, the authors assume that the distribution of the returns follows an infinite mixture of Normals via Dirichlet process mixture (DPM) models (see related works[24,25,26] among others). Such infinite mixtures turn out to be a very flexible modeling approach because they nest other parametric specifications for the error term. The infinite mixture of Normals can approximate other distributions, frequently used in financial time series context (see, eg, the works of Tokdar[27] and Mencía and Sentana,[28]) because of its "universal approximation property".[29] Finally, the use of the DPM models in financial time-series context is not restricted to density estimation only and can be used for other types of infinite mixtures. For example, Lau and So[30] considered infinite mixtures of autoregressive processes and showed how to perform cluster analysis and outlier detection using the DPM. Lau and So[31] highlighted that the mixture of time series models can capture structural changes and showed that mixing over infinite GARCH processes results into superior model performance for 10 index return series.

In the semiparametric SV models, even when the distribution of the returns is modeled in a flexible nonparametric manner, the volatility equation still maintains its simple AR(1) representation. Such model might have some limitations by not allowing for structural changes in the volatility process. If these changes are not accounted for, the persistence parameter in the volatility equation might be overestimated. Overestimation of the persistence parameter leads to incorrect conclusions about the predictability of the volatility.[32] Therefore, in this paper, we augment the semiparametric stochastic volatility model, similar to the one in the work of Delatola and Griffin,[21] to include Markov switching regimes in the volatility equation, resulting into a Bayesian semiparametric Markov switching SV (MSSV-DPM) model. Including shifts in the volatility regimes was first proposed by So et al.[33] Since then, fully parametric MSSV models have been rather popular in the financial time series context due to its superior performance as compared to the benchmark SV models. Kalimipalli and Susmel[34] considered two-factor SV model with regime switching and found that the estimated high volatility persistence is reduced when the regimes are incorporated in the model. Shibata and Watanabe[35] also found that the persistence parameter estimates drop as compared to those of the standard SV models. Moreover, for their data, the MSSV model performs better than the benchmark SV models. Similar findings are also present in the work of Vo,[32] who models oil price movements. Carvalho and Lopes[36] used an auxiliary particle filter (APF) to sequentially learn about states and parameters of the MSSV model and showed the predictive superiority of the MSSV model. For surveys on regime switching models, GARCH, and SV, refer to related works[37-39] among others.

In general, the estimation of SV-type models is rather complex, given the unobservable nature of the volatility. The Markov chain Monte Carlo (MCMC) is the standard approach in the Bayesian context, with the seminal work by Jacquier et al.[5] For a survey on Bayesian estimation of time-varying volatility models, see the works of Chen et al[37] and Virbickaite et al[40] among others. Even though MCMC methods are considered to be the gold standard among Bayesian estimation methods, they are computationally costly and inherently nonsequential.[41] A cost-efficient alternative to MCMC is sequential Monte Carlo (SMC) methods, also known as particle filters, that allow for on-line type inference by updating the posterior distribution as new observations arrive. By construction, stochastic volatility models are state-space models, naturally suggesting the use of particle filters for estimation. Moreover, the model proposed in this paper belongs to a class of models that have the availability of sufficient statistics of the parameters (see the work of Storvik[42]). This permits to track a low-dimensional set of sufficient statistics instead of a high-dimensional vector of parameters. The use of sufficient statistics has been shown to increase the efficiency of the algorithm by reducing the variance of the sampling weights (see the work of Carvalho et al[43]). In this paper, we make use of the particle learning (PL) approach, which is a particle-based method, firstly introduced by Carvalho et al.[43] For general introduction to PL and comparison with MCMC, see the works of Lopes and Polson[41] and Carvalho et al[43] among others. Warty et al[44] proposed a sequential estimation algorithm for the SV model with variance-gamma jumps in the returns. The algorithm, a hybrid between the APF and PL, is compared to the MCMC output and used for a real-data application. In a recent paper, Virbickaitė et al[45] have designed a PL algorithm for a semiparametric stochastic volatility model of Delatola and Griffin.[21] The authors conducted an extensive comparison with the MCMC estimation output and showed that both estimation methods present almost identical posterior distributions for model parameters, filtered volatility states, and the distribution of the error term. In this paper, we construct a PL algoritm similar to the one in the work of Virbickaitė et al[45] and augment it to include the Markov switching regimes.

The rest of this paper is structured as follows. Section 2 presents the linearized SV model with nonparametric errors and introduces a new MSSV-DPM model. Section 3 designs a PL algorithm for inference and prediction and presents a

simulated data example. Section 4 evaluates the performance of the proposed model by using real data. Finally, Section 5 concludes.

2 | MSSV-DPM MODEL

We start this section by reviewing a benchmark stochastic volatility model with Normal errors. We then relax the Normality assumption and present the semiparametric SV model, similar to the one seen in the work of Delatola and Griffin.[21] The innovation distribution is assumed to follow an infinite mixture of Gaussians via Dirichlet process mixture models, giving rise to a SV-DPM model. Finally, we augment the semiparametric SV model with Markov switching regimes in the volatility equation, resulting into a novel MSSV-DPM model.

Denote y_t as the de-meaned log returns that are modeled as $y_t = \exp\{h_t/2\}v_t$. The standard discrete SV model assumes the following dynamics for the volatility: $h_t = \alpha + \beta h_{t-1} + \tau\eta_t$. Here, β is the volatility persistence parameter such that $|\beta| < 1$ for the stationarity of the volatilities; v_t and η_t are uncorrelated error terms, such that $\eta_t \sim \mathcal{N}(0,1)$. The distribution of the v_t has zero mean and unit variance and can take many different forms: from the standard Normal to heavy-tailed Student-t and others (see related works[6,10,17,18] for example).

Kim et al[6] proposed a linearization of the standard SV model by defining $r_t = \log y_t^2$ and $\epsilon_t = \log v_t^2$, resulting into the following dynamic linear model:

$$r_t = h_t + \epsilon_t, \tag{1}$$

$$h_t = \alpha + \beta h_{t-1} + \tau\eta_t. \tag{2}$$

The distribution of ϵ_t is $\log \chi_1^2$ if v_t in the conventional SV model, described previously, is Normally distributed. Kim et al[6] and Omori et al[46] used carefully tuned finite mixtures of Normals to approximate the $\log \chi_1^2$ distribution and used a data augmentation argument to design fast MCMC schemes that jointly sample $\{h_1, \ldots, h_T\}$ based on the well-known forward filtering, ie, backward sampling algorithm of Carter and Kohn[47] and Fruhwirth-Schnatter.[48] However, if v_t is not Normally distributed, then approximations of Kim et al[6] and Omori et al[46] are not appropriate anymore.

2.1 | DPM errors

As mentioned in the introduction, it has been shown in multiple empirical studies that the distribution of the returns has heavier tails than permitted by the Normal distribution. If this is the case, the distribution of ϵ_t in (1) is not $\log \chi_1^2$ anymore. Delatola and Griffin[21,22] proposed to approximate the distribution of ϵ_t as an infinite mixture of Normals by relying on DPM models. Dirichlet process mixture models, firstly introduced by Lo,[26] have been widely used for modeling time-varying volatilities with univariate and multivariate SV and GARCH-type models (see other works[19-23,49-52]).

As seen in the work of Escobar and West,[53] the DPM model has the following representation:

$$f(\epsilon_t; G) = \int k(\epsilon_t; \theta_t) dG(\theta_t), \tag{3}$$

where k is some density kernel with parameter vector θ_t and the mixing distribution G has a Dirichlet Process prior, denoted here by $G \sim \mathcal{DP}(c, G_0(\theta; \varrho))$. Each observation ϵ_t comes from a kernel density $k(\cdot)$ with some parameters θ_t, following the mixing distribution G. The parameter c is called the concentration parameter and $G_0(\theta; \varrho)$ is called the base distribution with certain hyperparameters ϱ. The concentration parameter c can be seen as the prior belief about the number of clusters in the mixture. Small values of c assume a priori an infinite mixture model with only few components that have large weights. Meanwhile, large values of c assume a priori an infinite mixture model with many components and all the weights being very small. c is also called a precision parameter and indicates how close G is to the base distribution G_0, where larger c indicates that G is closer to G_0.

Gaussian kernel and conjugate base prior. One of the most popular DPM model variants in the financial time series context assumes a Gaussian kernel for $k(\epsilon_t; \theta_t)$ in (3), ie, $\epsilon_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$. Then, the conjugate base prior $G_0(\mu, \sigma^2; \varrho)$ is a Normal-inverse gamma prior, denoted here by $G_0 \sim \mathcal{NIG}(\mu, \sigma^2; m_0, V_0, a_0, a_0\sigma_0^2)$, such that $\mu|\sigma^2$ is Normal, ie, $\mathcal{N}(\mu; m_0, V_0\sigma^2)$, and σ^2 is inverse gamma, ie, $\mathcal{IG}(\sigma^2; a_0/2, a_0\sigma_0^2/2)$. Here, m_0, V_0, a_0, and $a_0\sigma_0^2$ are the hyperparameters in ϱ.

2.2 | Markov switching volatility

As mentioned in the introduction, the benchmark stochastic volatility model has certain limitations. In particular, it does not account for structural changes in the volatility process and, if such regime shifts are ignored, the persistence parameter is overestimated. In other words, the β parameter in (2) is very close to one and the volatility equation approaches the nonstationary process. In order to incorporate such changes in the regimes, So et al[33] introduced the MSSV model, where the log volatility equation is of the following form:

$$h_t = \alpha_{s_t} + \beta h_{t-1} + \tau\eta_t,\ \eta_t \sim \mathcal{N}(0,1). \tag{4}$$

Here, s_t are the regime variables following a two-state first order Markov process

$$p_{ij} = \mathrm{P}[s_t = j | s_{t-1} = i],\ \text{for } i,j = 0,1.$$

As seen in the work of Carvalho and Lopes,[36] it is necessary to introduce the following reparametrization for α_{s_t} to avoid identification issues:

$$\alpha_{s_t} = \gamma_0 + \gamma_1 \boldsymbol{I}\{s_t = 1\},\ \gamma_0 \in \Re \text{ and } \gamma_1 > 0.$$

Here, $\boldsymbol{I}\{s_t = 1\}$ is an indicator function that takes values equal to one if the volatility is in the high state ($s_t = 1$) and zero in the low state ($s_t = 0$). The transition matrix between the states 0 and 1 is defined as follows:

$$T = \begin{bmatrix} P(s_t = 0|s_{t-1} = 0) & P(s_t = 1|s_{t-1} = 0) \\ P(s_t = 0|s_{t-1} = 1) & P(s_t = 1|s_{t-1} = 1) \end{bmatrix} = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}. \tag{5}$$

There are quite a few papers that consider regime switching SV models in Bayesian context. Kalimipalli and Susmel[34] have proposed a two-factor SV model with regime switches and estimated it using Gibbs sampler. They found that the estimate of high volatility persistence is reduced when the regimes are incorporated in the model. Moreover, the authors compared the new model with other two alternative two-factor models, simple SV, and GARCH, and found that SV always outperforms GARCH, both in sample and out of sample. Shibata and Watanabe[35] designed an MCMC scheme to estimate the MSSV model and found that the persistence parameter estimates drop as compared to those of the standard SV models. Moreover, the MSSV model performs better than the benchmark SV models. Lopes and Carvalho[54] extended the SV model to multivariate case and presented a factor stochastic volatility (FSV) model with Markov switching jumps. They constructed a novel MCMC scheme for inference and found that the new model can capture market crashes in an instantaneous way, as opposed to the traditional FSV models. Carvalho and Lopes[36] have constructed an SMC filter by combining APF with the filter of Liu and West[55] to estimate an SV model with Markov switching regimes. They found that, in terms of predictions, the Markov switching SV specification outperforms a simple SV model. Abanto-Valle et al[56] investigated the relationship between stock return volatility and trading volume by using a MSSV specification and also found that the persistence parameter drops significantly after introducing the Markov switching jump. Finally, Chen et al[57] considered a Markov switching GARCH model, where the volatility states are determined by a hidden Markov chain, and the return distribution is assumed to follow a fat-tailed t. Their model could be seen as a GARCH-counterpart of the model proposed in here (only we allow for fat-tailed errors via DPM specification). In addition, the authors allowed for all volatility parameters to change across different states as well as the mean equation for the returns, allowing for more flexibility. Chen et al[57] showed that their proposed model is superior to standard approaches in VaR estimation.

Define $\boldsymbol{\Phi} = (\gamma_0, \gamma_1, \beta, \tau^2, p, q)$ as a set of parameters associated with the volatility equation, $\boldsymbol{\Omega} = \{(\mu, \sigma^2)^{(j)}\}_{j=1}^{\infty}$ as a set of parameters associated with the distribution of the error term, and $\boldsymbol{\Theta} = (\boldsymbol{\Phi}, \boldsymbol{\Omega})$ as a complete set of model parameters. Therefore, the complete MSSV-DPM model is a linearlized SV model in (1)-(2) with DPM errors in (3) that accommodates the regime-shifting structure in (4)-(5) and can be written as follows:

$$r_t | h_t, \boldsymbol{\Theta} \sim \frac{1}{c+t-1} \sum_{j=0}^{L^\star_{t-1}} n_{t-1,j} \mathcal{N}\left(r_t; \mu_j + h_t, \sigma_j^2\right), \tag{6}$$

$$h_t | h_{t-1}, \lambda_t, \boldsymbol{\Theta} \sim \mathcal{N}\left(h_t; \gamma_0 + \gamma_1\lambda_t + \beta h_{t-1}, \tau^2\right), \tag{7}$$

$$\lambda_t | \boldsymbol{\Theta} \sim \mathcal{BER}\left((1-p)^{1-\lambda_{t-1}} q^{\lambda_{t-1}}\right). \tag{8}$$

Here, $n_{t,j}$ is a number of observations assigned to the jth component at time t, $n_0 = c$, $L_t^\star$ is a number of nonempty components in the mixture at time t, ie, $L_t^\star$ is not fixed a priori and grows if new components are observed. Given this missing information, the mixture becomes finite, and the upper limit for the number of components is the number of observations. In practice, data tends to cluster, meaning that some observations come from the same component, thus

$L_t^\star \leq t$. Moreover, $\mathcal{BER}(\pi)$ denotes a Bernoulli distribution with parameter π and λ_t is a Bernoulli distributed state variable that takes value 1 if the volatility is in the high regime and zero otherwise. The newly proposed MSSV-DPM model contains the SV-DPM model as a special case when there is only one regime, ie, $p = 0$. It also nests the benchmark SV and MSSV models with Normal innovations when $L_t^\star = 1, \forall t = 1, \dots, T$.

3 | ESTIMATION AND SIMULATION STUDY

In this section, we present the algorithm to perform PL estimation for the novel MSSV-DPM model. By using simulated data, we show that the estimation algorithm is able to precisely estimate the parameters and the density of the squared log returns and filter the latent log volatilities and volatility regimes.

3.1 | PL for the MSSV-DPM model

In this section, we modify and augment the PL algorithm presented in the work of Virbickaitė et al[45] to include the Markov switching specification. Particle learning, as mentioned before, is one of several particle filters that allow to perform sequential state filtering and parameter learning. Moreover, PL, which was firstly introduced by Carvalho et al,[43] allows for sequential filtering, smoothing, and parameter learning by including state-sufficient statistics in a set of particles. For a more detailed explanation of PL with illustrations, refer to the works of Carvalho et al[43] and Lopes et al[58] among others. For comparison between PL and MCMC for the SV-DPM model, refer to the work of Virbickaitė et al,[45] and for comparison between APF+PL and MCMC for the SV model with variance-gamma jumps in the returns, refer to the work of Warty et al.[44]

The priors for model parameters and the initial states are chosen to be conditionally conjugate: $h_0 \sim \mathcal{N}(c_0, C_0)$, $\sigma^2 \sim \mathcal{IG}(a_0/2, a_0\sigma_0^2/2)$, $\mu|\sigma^2 \sim \mathcal{N}(m_0, V_0\sigma^2)$, $\tau^2 \sim \mathcal{IG}(b_0/2, b_0\tau_0^2/2)$, $\beta|\tau^2 \sim \mathcal{TN}_{(-1,1)}(m_\beta, V_\beta\tau^2)$, $\gamma_0 \sim \mathcal{N}(m_{\gamma_0}, V_{\gamma_0})$, $\gamma_1 \sim \mathcal{TN}_{(0,+\infty)}(m_{\gamma_1}, V_{\gamma_1})$, $p \sim \mathcal{B}(\alpha_p, \beta_p)$, and $q \sim \mathcal{B}(\alpha_q, \beta_q)$. Here, $\mathcal{TN}_{(a,b)}$ represents a Normal distribution, truncated at a and b, $\mathcal{B}$ is Beta distribution, and $c_0, C_0, a_0, a_0\sigma_0^2, m_0, V_0, b_0, b_0\tau_0^2, m_\beta, V_\beta, m_{\gamma_0}, V_{\gamma_0}, m_{\gamma_1}, V_{\gamma_1}, \alpha_p, \beta_p, \alpha_q$, and β_q are the fixed hyperparameters.

Call S_t a set of sufficient statistics, which contains all updated hyperparameters, necessary for the parameter simulation, as well as the three kinds of filtered state variables: the latent log volatilities h_t; the indicator variable k_t, which tells us to which mixture component the data point belongs to; and λ_t, the volatility regime indicator. The object we call particle at time t contains S_t. All necessary parameters can be easily simulated given the set of sufficient statistics. At each time t, we have a collection of N particles that provides approximations to the densities of interest. When this set of N particles passes from one time to another, t to $t + 1$, some of the particles disappear (the ones that are not representative with respect to the new data point), and some are repeated more than once to take their place (see the "Resampling" step later). Then, this resampled set of particles is modified to include the information from the new data point (see the "Sampling" and "Propagating" steps later).

In order to initiate the algorithm, initial parameter values are simulated from their corresponding priors. The initial set of sufficient statistics S_0 consists of $\{h_0^{(i)}\}_{i=1}^N$, which has been simulated from its prior; $\{k_t^{(i)}\}_{i=1}^N$, which at $t = 0$ are all set equal to 1 because, when the first observation arrives, it will belong to the first and only component, initial volatility regime $\{\lambda_0\}_{i=1}^N = 0$ and initial hyperparameters $\{a_0^{(i)}\}_{i=1}^N, \{a_0\sigma_0^{2(i)}\}_{i=1}^N, \dots$, which at time $t = 0$ are all the same across all particles. Then, for $t = 1 \dots, T$ and for each particle (i), the algorithm iterates through the following steps. For notation simplicity, we do not include the indicator (i) that refers to a single particle, where $(i) = 1, \dots, N$.

1. **Resampling**.
   Resample the $(i) = 1, \dots, N$ particles with weights proportional to the predictive density of the log squared returns $r_t = \log y_t^2$:

$$w^{(i)} \propto \frac{1}{c+t-1} \sum_{j=0}^{L_{t-1}^\star} n_j f_N\left(r_t; \gamma_0 + \gamma_1\lambda_{t-1} + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2\right).$$

   Here, $\Theta = (\gamma_0, \gamma_1, \beta, \tau^2, p, q, \mu_1, \dots, \mu_{L_{t-1}^\star}, \sigma_1^2, \dots, \sigma_{L_{t-1}^\star}^2)$ have been simulated at the end of the previous period. The resampled particles are denoted by a tilde above the particle, as in $\tilde{\Theta}$.

2. **Sampling.**

(a) Sample new states of the log volatilities λ_t:

$$\lambda_t|\tilde{\lambda}_{t-1},\tilde{h}_{t-1},\tilde{\Theta},r_t \sim \mathcal{BER}\left(\frac{z_2}{z_1+z_2}\right),$$

where

$$\begin{aligned} z_1 &= \left[\frac{1}{c+t-1}\sum_{j=1}^{\tilde{L}^\star_{t-1}}\tilde{n}_j f_N\left(r_t;\tilde{\gamma}_0+\tilde{\beta}\tilde{h}_{t-1}+\tilde{\mu}_j,\tilde{\tau}^2+\tilde{\sigma}_j^2\right)\right. \\ &\quad \left.+\frac{c}{c+t-1}f_N\left(r_t;\tilde{\gamma}_0+\tilde{\beta}\tilde{h}_{t-1}+\mu_0,\tilde{\tau}^2+\sigma_0^2\right)\right]\times \Pr\left(\lambda_t=0|\tilde{\lambda}_{t-1},\tilde{\Theta}\right), \\ z_2 &= \left[\frac{1}{c+t-1}\sum_{j=1}^{\tilde{L}^\star_{t-1}}\tilde{n}_j f_N\left(r_t;\tilde{\gamma}_0+\tilde{\gamma}_1\tilde{\lambda}_{t-1}+\tilde{\beta}\tilde{h}_{t-1}+\tilde{\mu}_j,\tilde{\tau}^2+\tilde{\sigma}_j^2\right)\right. \\ &\quad \left.+\frac{c}{c+t-1}f_N\left(r_t;\tilde{\gamma}_0+\tilde{\gamma}_1\tilde{\lambda}_{t-1}+\tilde{\beta}\tilde{h}_{t-1}+\mu_0,\tilde{\tau}^2+\sigma_0^2\right)\right]\times \Pr\left(\lambda_t=1|\tilde{\lambda}_{t-1},\tilde{\Theta}\right). \end{aligned} \tag{9}$$

Call $\tilde{\alpha}=\tilde{\gamma_0}+\tilde{\gamma_1}\lambda_t$.

(b) Sample new log volatilities h_t:

$$h_t|\tilde{h}_{t-1},\tilde{\Theta},\tilde{L}^\star_{t-1},\lambda_t,r_t \sim \sum_{j=0}^{\tilde{L}^\star_{t-1}}\frac{\tilde{n}_j}{c+t-1}\mathcal{N}(h_t;m_{hj},V_{hj}),$$

where

$$m_{hj}=\frac{\tilde{\tau}^2(r_t-\tilde{\mu}_j)+\tilde{\sigma}_j^2(\tilde{\alpha}+\tilde{\beta}\tilde{h}_{t-1})}{\tilde{\tau}^2+\tilde{\sigma}_j^2} \text{ and } V_{hj}=\frac{\tilde{\sigma}_j^2\tilde{\tau}^2}{\tilde{\sigma}_j^2+\tilde{\tau}^2}.$$

For each particle, we sample h_t from a mixture of $L^\star_{t-1}+1$ components with the corresponding weights from the previous period.

(c) Sample new indicators k_t from $\{1,\ldots,L^\star_{t-1}+1\}$, with weights proportional to

$$\tilde{n}_j f_N\left(r_t;\tilde{\alpha}+\tilde{\beta}\tilde{h}_{t-1}+\tilde{\mu}_j,\tilde{\tau}^2+\tilde{\sigma}_j^2\right),\ j=1,\ldots,L^\star_{t-1},$$

where $\tilde{n}_{L^\star_{t-1}+1}=c$ and $\sigma^2_{L^\star_{t-1}+1}=\sigma_0^2$. If $k_t\le L^\star_{t-1}$, $n_{k_t}=\tilde{n}_{k_t}+1$ and $L^\star_t=L^\star_{t-1}$, otherwise, $L^\star_t=L^\star_{t-1}+1$ and $n_{k_t}=1$. Note that the method described here of how to sequentially learn the number of components L^*_t can be seen as an online adaptation of the weighted Chinese restaurant process of Lo.[59] The exact same weighting approach was considered by Lau and So,[30] for example, where the weights are proportional to the cluster sizes adjusted by the predictive density of the observation equation, the only difference being that the authors use MCMC methods for inference, which also allows to use the *entire* sample information. In the SMC setting, however, it is impossible to incorporate information from the so-called "future" seating arrangements because of the sequential nature of the algorithm, and the observations, once assigned to a certain cluster, cannot be reseated.

3. **Propagating sufficient statistics and learning Θ.**

(c.1) Sample γ_0 from $\mathcal{N}(\gamma_0;m^\star_{\gamma_0},V^\star_{\gamma_0})$, where

$$m^\star_{\gamma_0}=\frac{\tilde{m}_{\gamma_0}\tilde{\tau}^2+\tilde{V}_{\gamma_0}\left(h_t-(\tilde{\gamma}_1\lambda_t+\tilde{\beta}\tilde{h}_{t-1})\right)}{\tilde{\tau}^2+\tilde{V}_{\gamma_0}} \text{ and } V^\star_{\gamma_0}=\frac{\tilde{\tau}^2\tilde{V}_{\gamma_0}}{\tilde{\tau}^2+\tilde{V}_{\gamma_0}}.$$

(c.2) Sample γ_1 from $\mathcal{TN}_{(0,+\infty)}(\gamma_1;m^\star_{\gamma_1},V^\star_{\gamma_1})$, where

$$m^\star_{\gamma_1}=\frac{\tilde{m}_{\gamma_1}\tilde{\tau}^2+\tilde{V}_{\gamma_1}\lambda_t\left(h_t-(\gamma_0+\tilde{\beta}\tilde{h}_{t-1})\right)}{\tilde{V}_{\gamma_1}\lambda_t+\tilde{\tau}^2} \text{ and } V^\star_{\gamma_1}=\frac{\tilde{\tau}^2\tilde{V}_{\gamma_1}}{\tilde{\tau}^2+\lambda_t\tilde{V}_{\gamma_1}}.$$

Call $\alpha = \gamma_0 + \gamma_1 \lambda_t$.

(c.3) Sample τ^2 from $\mathcal{IG}(\tau^2; b_0^\star/2, b_0^\star \tau_0^{2\star}/2)$, where

$$b_0^\star = \tilde{b}_0 + 1 \text{ and } b_0^\star \tau_0^{2\star} = \tilde{b}_0 \tilde{\tau}_0^2 + \frac{\left(\tilde{m}_\beta \tilde{h}_{t-1} - (h_t - \alpha)\right)^2}{1 + \tilde{V}_\beta \tilde{h}_{t-1}^2}.$$

(c.4) Sample β from $\mathcal{TN}_{(-1,1)}(\beta; m_\beta^\star, V_\beta^\star \tau^2)$, where

$$m_\beta^\star = \frac{\tilde{m}_\beta + \tilde{V}_\beta \tilde{h}_{t-1}(h_t - \alpha)}{1 + \tilde{V}_\beta \tilde{h}_{t-1}^2} \text{ and } V_\beta^\star = \frac{\tilde{V}_\beta}{1 + \tilde{V}_\beta \tilde{h}_{t-1}^2}.$$

(c.5) Sample p from $\mathcal{B}(p; \alpha_p^\star, \beta_p^\star)$, where

$$\alpha_p^\star = \alpha_p + 1 \text{ if } \lambda_t = 0 | \lambda_{t-1} = 0 \text{ and } \beta_p^\star = \beta_p + 1 \text{ if } \lambda_t = 1 | \lambda_{t-1} = 0.$$

(c.6) Sample q from $\mathcal{B}(q; \alpha_q^\star, \beta_q^\star)$, where

$$\alpha_q^\star = \alpha_q + 1 \text{ if } \lambda_t = 1 | \lambda_{t-1} = 1 \text{ and } \beta_q^\star = \beta_q + 1 \text{ if } \lambda_t = 0 | \lambda_{t-1} = 1.$$

(c.7) Sample $\sigma_{k_t=j}$ only for that component j, where the data point at time t is assigned to, ie, $k_t = j$, from $\mathcal{IG}(\sigma_{k_t=j}^2; a_0^\star/2, a_0^\star \sigma^{2\star}/2)$, where

$$a_0^\star = \tilde{a}_0 + 1 \text{ and } a_0^\star \sigma_0^{2\star} = \tilde{a}_0 \tilde{\sigma}_0^2 + \frac{(r_t - h_t - \tilde{m}_0)^2}{1 + \tilde{V}_0}.$$

(c.8) Sample $\mu_{k_t=j}$ only for that component j, where the data point at time t is assigned to, ie, $k_t = j$, from $\mathcal{N}(\mu_{k_t=j}; m_0^\star, V_0^\star \sigma_{k_t=j}^2)$, where

$$m_0^\star = \frac{\tilde{m}_0 + \tilde{V}_0(r_t - h_t)}{1 + \tilde{V}_0} \text{ and } V_0^\star = \frac{\tilde{V}_0}{1 + \tilde{V}_0}.$$

Parts of the derivations of the equations are available in the Appendix at the end of the manuscript, whereas the rest are available in the work of Virbickaitė et al.[45]

3.2 | Simulation study

In order to asses the estimation accuracy of the PL algorithm for the proposed model, we use a simulated data set of $T = 2000$ observations with the following parameters: $\gamma_0 = -0.06$, $\gamma_1 = 0.15$, $\beta = 0.92$, $\tau^2 = 0.05$, $p = 0.996$, and $q = 0.996$. The error term for the returns follows a standard Normal distribution $v_t \sim \mathcal{N}(0, 1)$, thus the true DGP for the linearlized model is $\epsilon_t \sim \log \chi_1^2$. The hyperparameters are $c_0 = 0$, $C_0 = 0.1$, $m_\beta = 0.95$, $V_\beta = 0.1$, $b_0 = 4$, $b_0 \tau_0^2 = 0.2$, $a_0 = 5$, $a_0 \sigma_0^2 = 15$, $m_0 = -1.27$ (this specific value is chosen because the mean of the $\log \chi_1^2$ distribution is equal to -1.27), $V_0 = 0.1$, $m_{\gamma_0} = 0$, $V_{\gamma_0} = 1$, $m_{\gamma_1} = 0$, $V_{\gamma_1} = 0.1$, $\alpha_p = 3$, $\beta_p = 0.1$, $\alpha_q = 3$, and $\beta_q = 0.1$. The concentration parameter c is set to be equal to 1, as in the work of Carvalho et al.[60] Using the simulated data, we fit the MSSV-DPM model using PL, number of particles $N = 300$k. All codes were written in R.[†]

Figure 1 top graph draws the simulated log returns y_t. The middle graph represents the true realization of the log volatility (in black) and the mean estimated filtered log volatility (in gray). The 95% credible intervals (CIs) for the estimated filtered log volatility almost always capture the true realization of the log volatility (not reported). The bottom graph of the same figure draws the mean probability of being in a state one ($s_t = 1$), compared to the true state. As seen from the figure, PL takes some time to learn because, at first, it is not able to distinguish the regimes well. However, around

[†]Large data sets with a large number of particles were estimated on a multiprocessor server, so the running times are not reported. However, for the reference, we did a test run of four parallel MC chains with 100 000 particles each for a simulated data set of 2000 observations on a personal computer. The specifications are as follows: 8-core 3.00 GHz AMD Ryzen processor with 16 GB of RAM. Each run took between 2 hours 45 minutes and 2 hours 47 minutes.

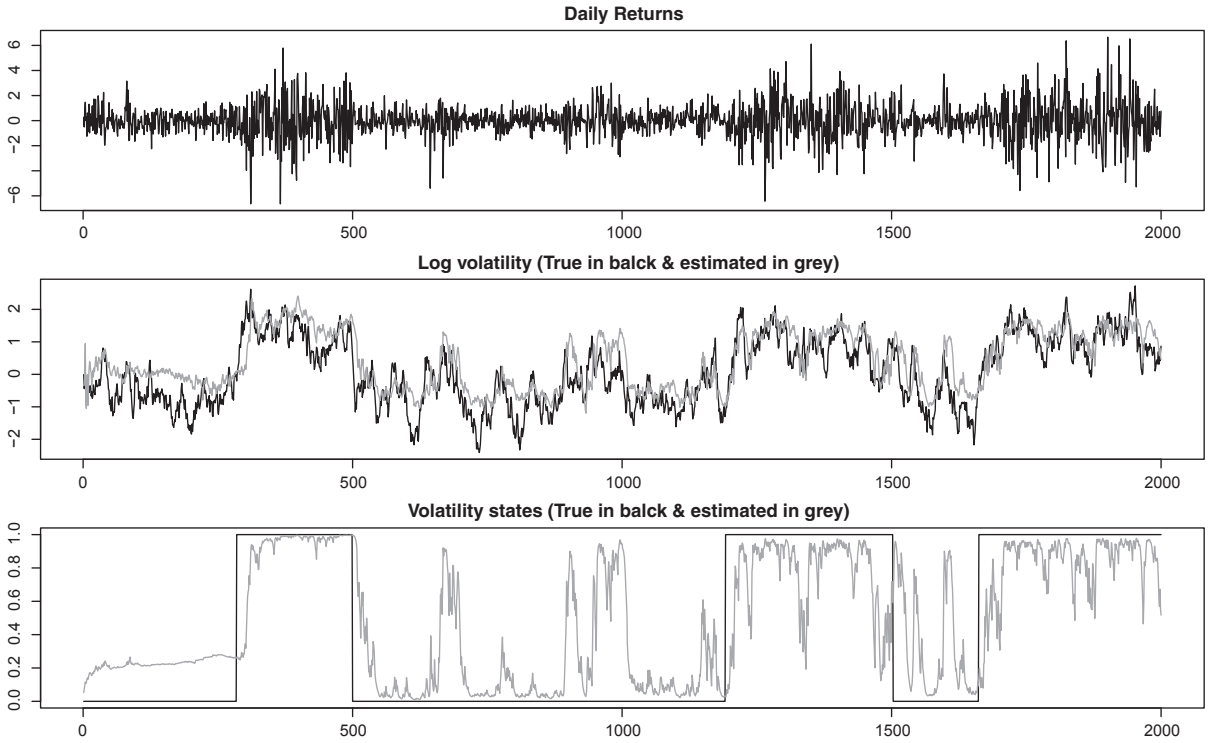

FIGURE 1 Simulated data: daily log returns (top), true and estimated log volatilities (middle), true and estimated volatility regimes (bottom)

observation 1000, the algorithm is able to correctly identify the regimes with the overall regime misclassification rate equal to 13%. Figure 2 draws the sequential estimation of the model parameters and their 95% CIs. The parameter estimates seem reasonable and very close to their true values. It takes some time to learn the true value of parameter q because there is no information available about it before the regime switch, which happens around time $t = 300$. The estimation of the parameter τ^2 is the least precise due to the fact that this parameter represents the volatility of the volatility and is notoriously difficult to estimate. Finally, Figure 3 draws the estimated nonparametric density of the log squared error term, which is almost identical to the true DGP, which is log χ_1^2 distributed. We have repeated the same exercise four times to get some intuition about the size of Monte Carlo error. For the estimation results across four independent runs, see Table 1. We can see that the 95% CIs at time T are rather similar across all four runs, maybe except parameter γ_1 because it can be estimated only when volatility is in high regime. The differences arise because of the Monte Carlo error. Increasing the number of particles would result into almost identical estimation paths.

Overall, the obtained estimation results seem quite reasonable and PL is able to correctly identify the hidden volatility regimes, filter log volatilities, estimate the density of the errors, and the parameters in an efficient sequential manner. The model diagnostics can be done by plotting the nonparametrically estimated density vs. the actually observed log squared returns, as seen in Figure 3. Furthermore, one can also inspect the estimated Markov switching matrix to see if it is diagonal or not to decide if Markov switching assumption is valid. Alternatively, one could also perform the residuals analysis, among many other model diagnostic checks.

As for the choice of the size of the particle approximation of the densities of interest, there are no well-established rules in deciding the number of particles. However, one should choose the number of particles such that the estimated parameter paths do not exhibit particle degeneracy, ie, a well-known shortcoming of particle filters. Virbickaitė et al[45] summarized the main weaknesses of particle filters, with the main disadvantage being an ever-decreasing set of atoms in the particle approximation of the density of interest. As noted by Chopin et al,[61] increasing the number of observations will lead to degenerating paths, unless the number of particles is being increased simultaneously. Therefore, the use of PL or any particle-based filter in general is advantageous only if one is interested in fast one-step-ahead predictions. However, once the number of observations has increased, one should consider restarting the filter at a later time t with a smaller number of observations, or anticipate the large sample size and employ more particles (which would slow down the estimation because the sizes of matrices that need to be carried from one time to another increase dramatically).

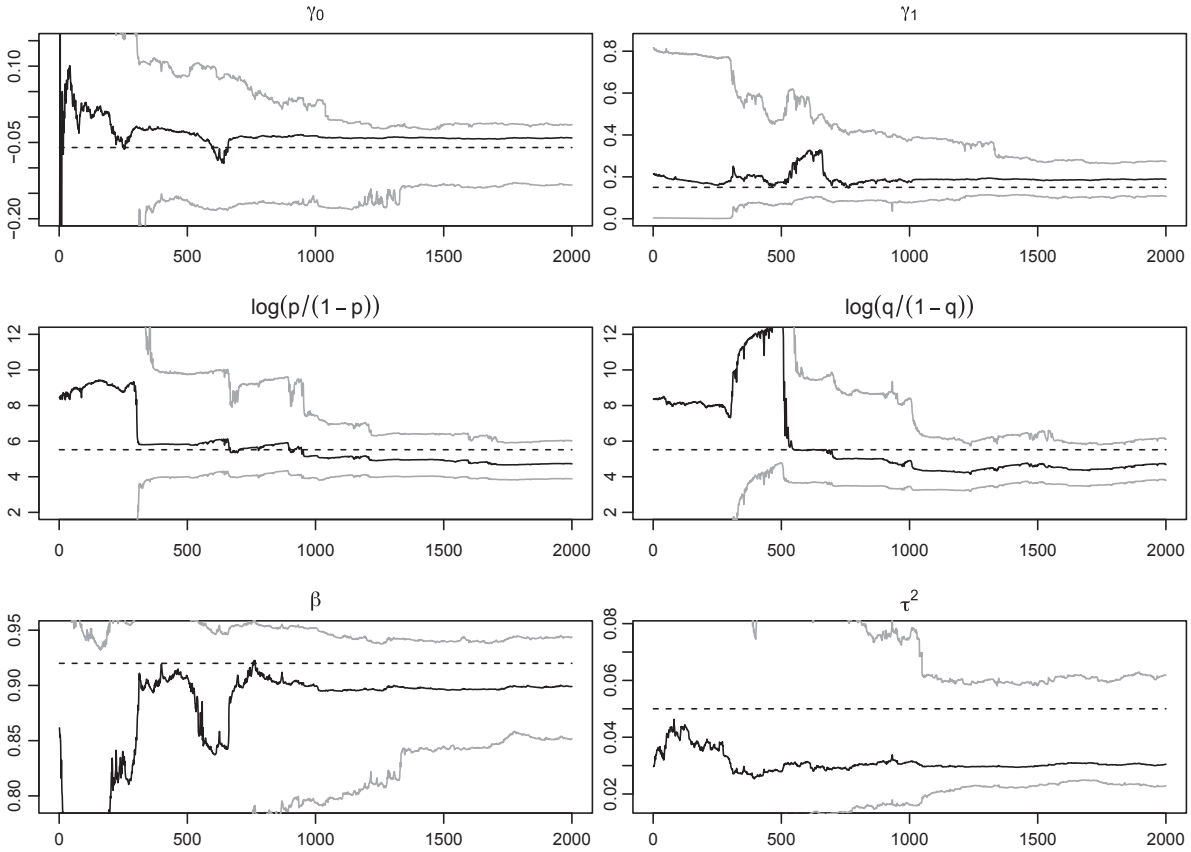

FIGURE 2 Simulated data: sequential estimation of the model parameters, their 95% credible intervals and their true values

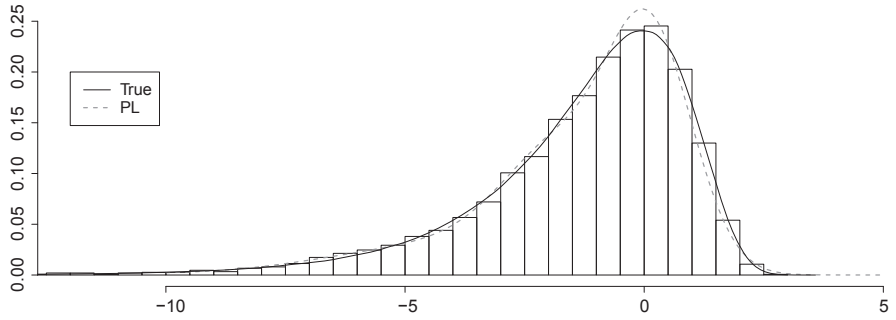

FIGURE 3 Simulated data: MSSV-DPM estimated nonparametric density of the log squared error term compared to the true DGP of $\log \chi_1^2$. DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; PL, particle learning

TABLE 1 95% credible intervals at time T for the estimated parameters of the MSSV-DPM model using the simulated data across four different runs

Parameter **True Value**	γ_0 **−0.06**	γ_1 **0.15**	p **0.996**	q **0.996**	β **0.92**	τ^2 **0.05**
Run 1	−0.090, −0.061	0.107, 0.178	0.987, 0.999	0.985, 0.999	0.907, 0.944	0.026, 0.044
Run 2	−0.134, −0.015	0.107, 0.273	0.980, 0.998	0.978, 0.998	0.852, 0.944	0.023, 0.062
Run 3	−0.119, −0.012	0.125, 0.247	0.983, 0.999	0.983, 0.999	0.864, 0.932	0.012, 0.060
Run 4	−0.107, −0.065	0.113, 0.180	0.984, 0.999	0.983, 0.999	0.900, 0.936	0.026, 0.042

Abbreviations: DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility.

TABLE 2 Descriptive statistics for S&P500, Ford, and natural gas data

	S&P500	Ford	Gas
Mean	0.0223	0.0182	0.0104
Median	0.0690	−0.0778	0.0668
St.dev.	1.2752	2.8026	4.4554
Skewness	−0.2237	−0.0220	0.7370
Kurtosis	10.4789	15.8981	28.3024
T	4447	4329	4193

Another well-documented issue associated with mixture models is the "label switching" phenomenon, which occurs when mixture parameters are unidentified. There are multiple (infinite) possible permutations of the mixture parameters, leading to the same density (see the work of Frühwirth-Schnatter[62]). However, label switching in MCMC context is different than in the SMC setting. In MCMC, the same observation at time t can be assigned to a different cluster label than in the previous sweep of the sampler. On the other hand, in SMC (and PL), once the observation at time t is assigned to a certain cluster, it is never reassigned because of the sequential nature of the sampler. However, there is no guarantee that an observation at time t is assigned to the same cluster j across all particles. In addition, if it is assigned to a different cluster, the associated cluster parameters are estimated to be different over the particles at time t. Therefore, as noted in the work of Jensen et al,[19] DPM clusters cannot be used to determine different states of the economy, for example. Nonetheless, the objective of using DPM is to model the unknown distribution of the log-squared error term, hence label switching does not present a problem because any permutation still leads to the same density.

4 | REAL DATA APPLICATION

In this section, we present a real data application using log returns of three financial assets: S&P500 index, Ford company, and a commodity, ie, natural gas. The S&P500 and Ford prices are from January 2, 1997, till September 9, 2014, and Henry Hub natural gas spot prices (dollars per million btu) are from January 5, 1997, till September 9, 2014. All data are obtained from the Datastream database. The descriptive statistics are in Table 2 and the descriptive graphs are in Figure 4.

Next, using the de-meaned data, we fit two semiparametric models, SV-DPM, and the newly introduced MSSV-DPM. The hyperparameters for the priors are the same as in the simulation study in Section 3.2. Note that the SV-DPM model is a restricted version of the MSSV-DPM model, where the probability of staying in the same regime is set equal to one, $p = 1$, therefore, neither q nor γ_1 nor λ_t are possible to estimate. Moreover, as mentioned before, both models nest a benchmark SV model with Normal errors as a special case. The number of particles is set to $N = 500$k because the number of observations is larger than in the simulation study (4000+).

To compare the performance of the models, we use the average log predictive score (LPS) and average log predictive tail score (LPTS_α). LPTS_α considers the predictive performance only in the upper $100\alpha\%$ of the empirical distribution of the squared log returns. LPTS_α was also employed by Delatola and Griffin.[21] As the authors pointed out, the LPTS_α is not a proper scoring rule, however, it can be very helpful in understanding how the model performs in the tails. The LPS is defined as follows:

$$\text{LPS} = -\frac{1}{T}\sum_{t=1}^{T} \log p(r_t|r^{t-1}),$$

and LPTS_α is defined as:

$$\text{LPTS}_\alpha = -\frac{1}{\sum_{t=1}^{T} \boldsymbol{I}\{r_t > z_\alpha\}} \sum_{t=1}^{T} \boldsymbol{I}\{r_t > z_\alpha\} \log p(r_t|r^{t-1}),$$

where z_α is the upper 100α percentile of the empirical distribution of r_t. Note that smaller values of the LPS and LPTS_α correspond to a better model.

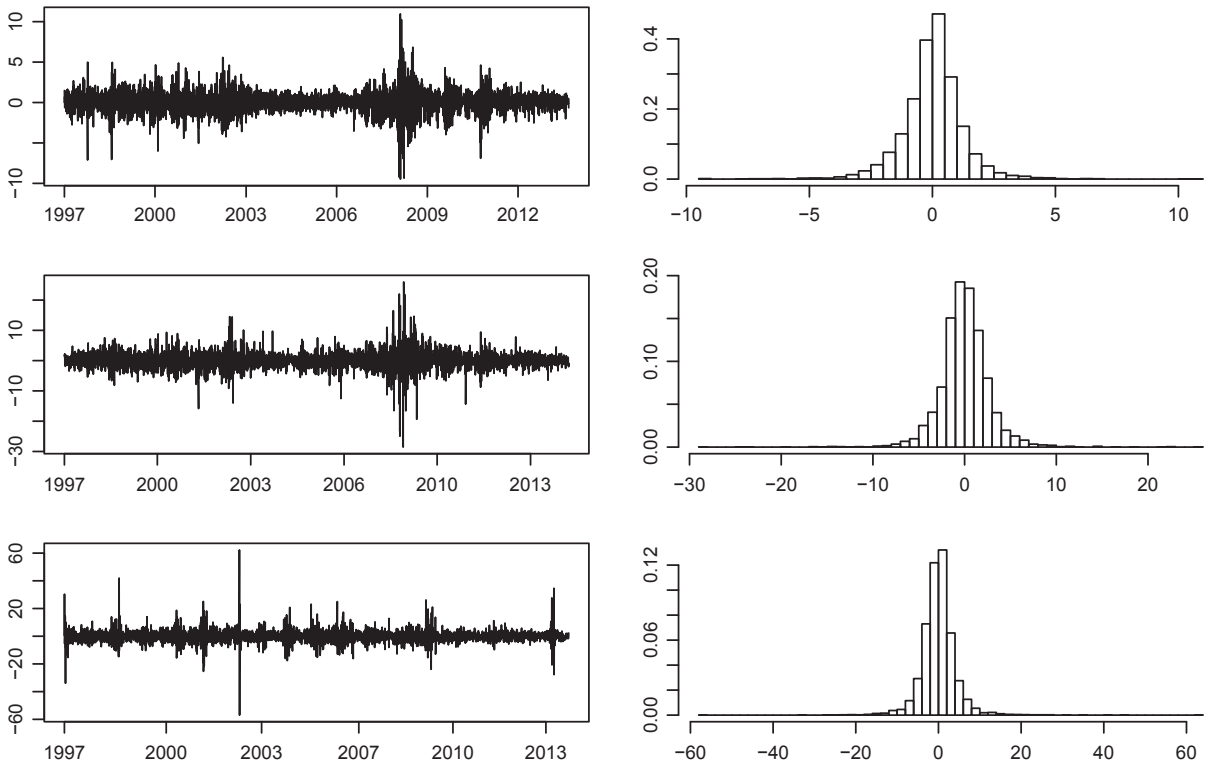

FIGURE 4 Daily log returns (in percentage) and histograms for the S&P500 (top), Ford (middle), and natural gas (bottom) data

The log predictive densities are straightforward to obtain with the PL algorithm because they are a by-product of the estimation procedure. For each $t = 1, \ldots, T$, the log predictive densities are calculated as follows:

$$\log p(r_t|r^{t-1}) = \frac{1}{N}\sum_{i=1}^{N} \log p\left(r_t|(\Theta, h_t, k_t, \lambda_t)^{(i)}\right). \tag{10}$$

Differently than in MCMC setting, there is no need to fix a certain $\hat{\Theta}$ for the calculation of the LPS and LPTS_α, and we can account for the parameter and state uncertainty by using the approximation in (10). Accounting for parameter and state uncertainty in MCMC setting at each time t without fixing certain $\hat{\Theta}$ would be prohibitively costly. The LPS and LPTS_α report the average predictive performance for the entire distribution and the tails of the distribution, respectively. Comparing these scores is not straightforward because there is no established decision rule to decide whether the difference in the scores between the two models is statistically significant.

Therefore, we also report the cumulative LPS, which can be seen as a log predictive Bayes factor (BF). Bayes factor allows for consistent model comparison even for nonnested models, it contains rewards for model fit, accounts for coherency

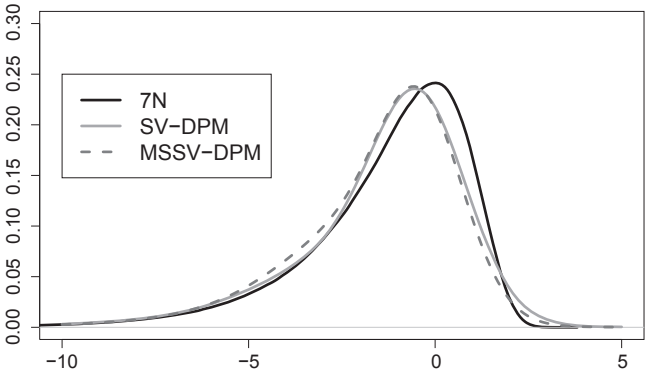

FIGURE 5 Estimated densities for the log squared error term for the SV-DPM and MSSV-DPM models, as compared to the approximation of 7 Normals. DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic volatility

between the prior and the information arising from the data, as well as rewards parsimony (see the work of Koop[63]). The BF between model $\mathcal{M}_1$ and model $\mathcal{M}_2$ is defined as $BF_{12} = p(D|\mathcal{M}_1)/p(D|\mathcal{M}_2)$ (see the work of Kass and Raftery[64]). Here, $p(D|\mathcal{M})$ is the marginal likelihood for data D given a certain model $\mathcal{M}$. For a predictive BF, this marginal likelihood is nothing else but the log predictive density in (10). Then, the difference between such cumulative log predictive densities is a log predictive BF (LPBF). Kass and Raftery[64] also provided a scale for the strength of preference of one model against another, and if the $2 \times \text{LPBF} > 10$, the evidence in favor of one model against another is very strong.

Next, we present the estimation results for the S&P500 index. Figure 5 draws the estimated densities for the error term for the SV-DPM and the MSSV-DPM models as compared to the frequently used mixture of 7 Normals of Kim et al,[6] as an approximation for $\log \chi_1^2$. If the distribution term of the returns was Normal, then the estimated distribution would be very similar to the mixture of 7 Normals. In our estimation results, the SV-DPM and MSSV-DPM models estimate the distribution term that is different from the 7-mixture approximation, therefore, we can conclude that the distribution term of the returns in the nonlinearlized model is anything but Normal. As a result, the assumption of Normality is restrictive and, for this data set, would be inappropriate. As seen in Figure 6, the filtered log volatilities and volatilities for both models are very similar (second and third graphs). The differences arise during the financial crisis period, where the MSSV-DPM model estimates higher volatility. This is one of the main advantages of the Markov switching specification because it allows for high volatility via the shift in the regime and not by artificial inflation of the volatility persistence parameter β. The filtered volatility regimes, ie, the mean probability of being in a state one ($s_t = 1$) for the MSSV-DPM model, are in the bottom part of Figure 6. The volatility of the S&P500 index is in the high regime during the years 1999-2001 (could be attributed to the dot-com bubble) and 2008-2010 (financial crisis). Table 3 presents the estimated parameter means and

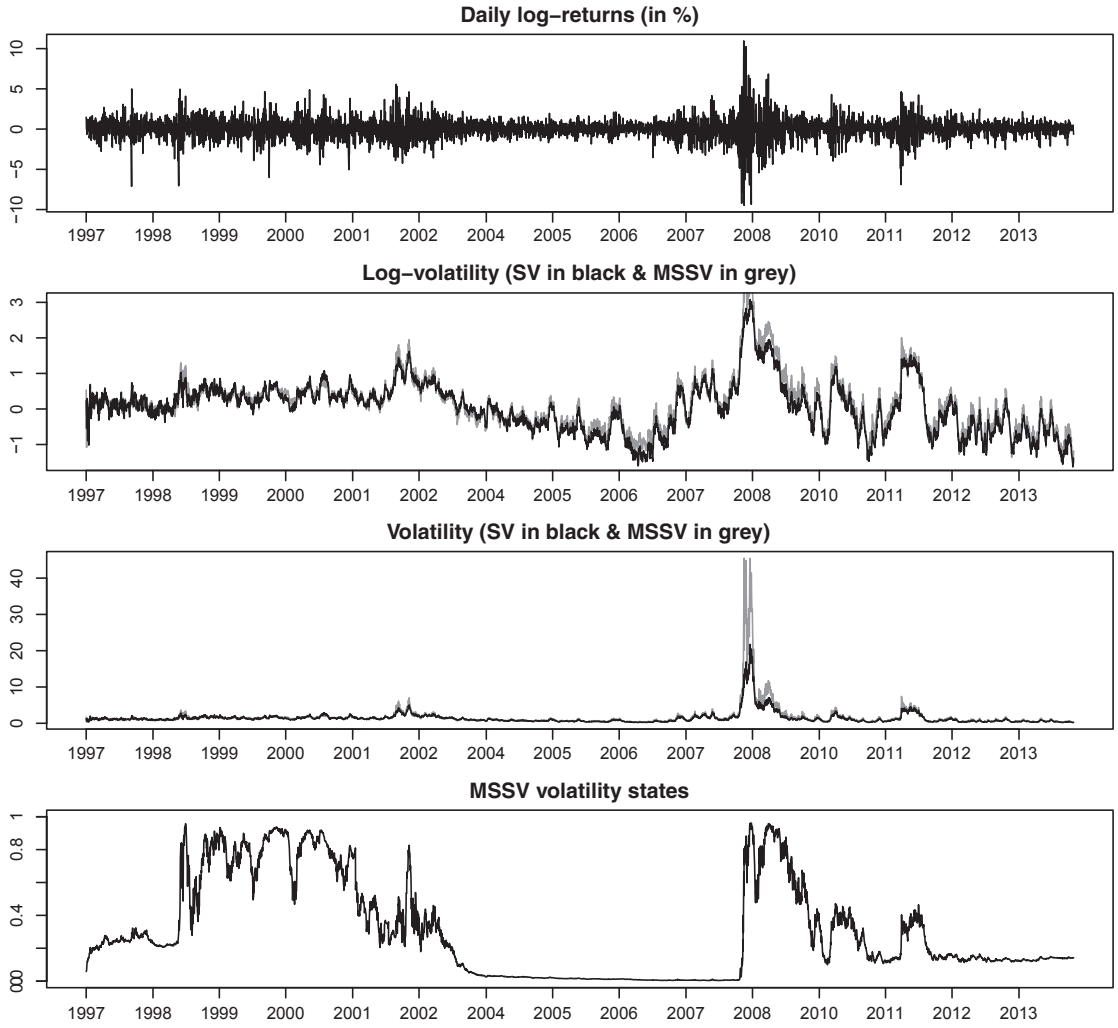

FIGURE 6 Filtered volatilities and volatility states (the mean probability of being in a state one $s_t = 1$) for S&P500 data for SV-DPM and MSSV-DPM models. DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic volatility

TABLE 3 Parameter estimation for SV-DPM and MSSV-DPM models for S&P500 data at time T

	SV-DPM		MSSV-DPM	
	Mean	**95% CI**	**Mean**	**95% CI**
α	0.0019	(−0.002, 0.0061)	-	-
β	0.9887	(0.984, 0.9932)	0.9712	(0.9592, 0.9822)
τ^2	0.016	(0.0144, 0.0173)	0.0319	(0.0293, 0.0351)
γ_0	-	-	0.0074	(6e-04, 0.0136)
γ_1	-	-	0.0737	(6e-04, 0.5593)
p	-	-	0.9998	(0.9987, 1)
q	-	-	0.9971	(0.8617, 1)

Abbreviations: CI, credible interval; DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic volatility.

TABLE 4 LPS and LPTS for SV-DPM and MSSV-DPM for S&P500 data ($T = 4447$)

	SV-DPM	**MSSV-DPM**	**Difference**	**2 × LPBF**
LPS	2.1956	2.1991	−0.0036	−32.0184
$\text{LPTS}_{0.10}$	2.5953	2.5527	0.0426	37.8884
$\text{LPTS}_{0.05}$	2.8400	2.7826	0.0574	25.5258
$\text{LPTS}_{0.01}$	3.3949	3.2398	0.1550	13.7857

Abbreviations: DPM, Dirichlet process mixture; LPBF, log predictive Bayes factor; LPS, log predictive score; LPTS, log predictive tail score; MSSV, Markov switching stochastic volatility; SV, stochastic volatility.

TABLE 5 Parameter estimation for SV-DPM and MSSV-DPM models for Ford data at time T

	SV-DPM		MSSV-DPM	
	Mean	**95% CI**	**Mean**	**95% CI**
α	0.0707	(0.0641, 0.0774)	-	-
β	0.9514	(0.9461, 0.9563)	0.908	(0.9041, 0.9119)
τ^2	0.0432	(0.0339, 0.0487)	0.0395	(0.0374, 0.0417)
γ_0	-	-	0.1287	(0.1223, 0.1351)
γ_1	-	-	0.139	(0.113, 0.167)
p	-	-	0.9994	(0.9981, 0.9999)
q	-	-	0.9949	(0.9847, 0.9992)

Abbreviations: CI, credible interval; DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic volatility.

95% CIs. The volatility persistence parameter is significantly larger for the SV-DPM model, ie, the estimated 95% CIs do not overlap. This result is in line with the findings present in other papers (see related works[32-35] among others).

Table 4 presents the LPS, LPTS_α, and LPBF for the S&P500 data. As mentioned before, the LPS measures the average predictive model performance for the entire distribution of the log squared returns and LPTS_α only for the tails. Even though the averages are virtually indistinguishable, the cumulative differences, ie, LPBFs, are significantly different. In particular, for the S&P500 data, the SV-DPM model performs better if we consider the LPS (ie, the whole predictive density), but the results change if we consider only the tails of the predictive distribution, where the newly introduced MSSV-DPM model provides significantly better performance.

Similar results can be seen in the estimation of the other two data sets (see Tables 5, 6, 7, and 8 and Figures 7 and 8). For Ford data, the estimated mean of the persistence parameter drops from 0.9514 to 0.908 and the 95% CIs do not overlap. For the natural gas data, the SV-DPM model estimates the persistence parameter to be 0.8889 as compared to the MSSV-DPM

TABLE 6 LPS and LPTS for SV-DPM and MSSV-DPM for Ford data ($T = 4329$)

	SV-DPM	MSSV-DPM	Difference	2× LPBF
LPS	2.0783	2.0756	0.0027	23.3766
$LPTS_{0.10}$	2.8134	2.7260	0.0873	75.5843
$LPTS_{0.05}$	3.1713	3.0334	0.1379	59.6969
$LPTS_{0.01}$	4.2408	3.9699	0.2709	23.4545

Abbreviations: DPM, Dirichlet process mixture; LPBF, log predictive Bayes factor; LPS, log predictive score; LPTS, log predictive tail score; MSSV, Markov switching stochastic volatility; SV, stochastic volatility.

TABLE 7 Parameter estimation for SV-DPM and MSSV-DPM models for gas data at time T

	SV-DPM		MSSV-DPM	
	Mean	95% CI	Mean	95% CI
α	0.2262	(0.2125, 0.2408)	-	-
β	0.8889	(0.8812, 0.8961)	0.8137	(0.807, 0.8206)
τ^2	0.1219	(0.1129, 0.1458)	0.1338	(0.1245, 0.1492)
γ_0	-	-	0.4161	(0.397, 0.4368)
γ_1	-	-	0.2758	(0.2255, 0.3205)
p	-	-	0.9951	(0.9907, 0.9976)
q	-	-	0.9802	(0.9655, 0.9904)

Abbreviations: CI, credible interval; DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic votality.

TABLE 8 LPS and LPTS for SV-DPM and MSSV-DPM for Gas data ($T = 4193$)

	SV-DPM	MSSV-DPM	difference	2× LPBF
LPS	2.1592	2.1529	0.0063	52.8318
$LPTS_{0.10}$	2.8845	2.7875	0.0970	81.3442
$LPTS_{0.05}$	3.2812	3.1095	0.1717	71.9938
$LPTS_{0.01}$	4.5151	4.2300	0.2851	23.9085

Abbreviations: DPM, Dirichlet process mixture; LPBF, log predictive Bayes factor; LPS, log predictive score; LPTS, log predictive tail score; MSSV, Markov switching stochastic volatility; SV, stochastic volatility.

model estimate of 0.8137, and the 95% CIs do not overlap either. The average LPS and $LPTS_\alpha$ for both models are very similar, however, the LPBFs favor the MSSV-DPM model in the entire distribution and in the tails for both data sets (see Tables 6 and 8). Figures 7 and 8 draw the estimated volatilities and volatility states for Ford and natural gas data. The Ford returns exhibit three periods of volatility increases. The increase in the years 1998-1999 might be due to a series of lawsuits against the company, which later were dismissed. The volatility increase in 2002-2003 might be due to the early 2000s economic recession in the US; meanwhile, the volatility increase in the 2008-2009 is clearly due to the global financial crisis. Natural gas data exhibits a very clear periodical pattern of shifts in the volatility regimes due to seasonal-like behavior of the returns. The increase in volatility coincides approximately with the coldest months of the year in the northern hemisphere, which is marked by increase demand in natural gas and possible increase in price uncertainty. Therefore, for gas data, the volatility states could be interpreted as seasonality-related states, which explains why the states are much less persistent compared to the other assets. Interestingly, during the global financial crisis, the natural gas volatility does not exhibit a long-lasting shift in the regime.

It is important to mention that Figures 6 to 8 present only filtered but not smoothed volatilities and volatility state estimates. In general, the purpose of the analysis plays essential role. If one is interested in understanding the historical behavior of the series and the effects of, say, economic factors on the changes in volatility, it is important to consider the information from the entire sequence, as is done by MCMC. In the SMC setting, this can be achieved by performing the

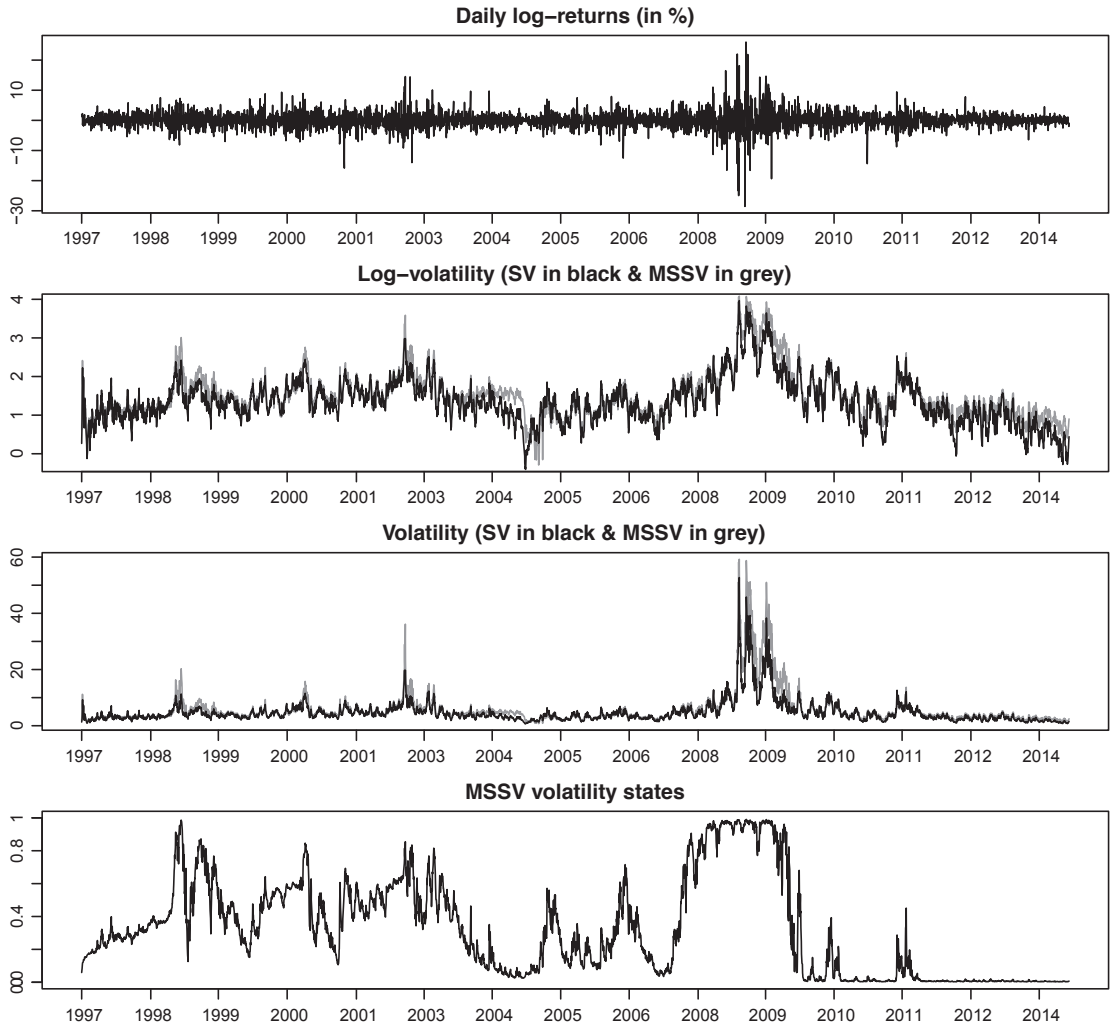

FIGURE 7 Filtered volatilities and volatility states (the mean probability of being in a state one $s_t = 1$) for Ford data for SV-DPM and MSSV-DPM models. DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic volatility

backwards smoothing procedure, where PL provides a procedure to perform direct backwards smoothing (see the work of Carvalho et al[43]). If one is interested in the prediction of the volatility in the next period $t + 1$, backwards smoothing does not apply. As noted in the work of Lopes et al,[58] in most models estimated, using PL smoothing can effectively be performed after the estimation.

To conclude, in majority of the cases, the newly proposed MSSV-DPM model outperforms the SV-DPM model in terms of one-step-ahead prediction for the entire distribution (except for the S&P500 data) and for the tails (for all three data sets). Moreover, including the regime shifts in the mean of the volatility equation reduces the value of the estimated persistence parameter. As noted in the work of Vo,[32] ignoring the shifts in the regimes gives the impression that the volatility is highly persistent, therefore, highly predictable, which is not the case. The half-life of a volatility shock, defined as "the time it takes for a shock to decay half of its initial value" (see the work of Vo[32]) for the S&P500 data drops from 61 days to 24 days; for Ford data, from 14 to 7 days; and for natural gas data, from 6 days to 3 days.

Improved model performance, especially in the tails, has potential implications for estimating the tail risk, such as value-at-risk (VaR), expected shortfall, among others. There is extensive evidence in the literature that includes regime switches in the volatility equation, coupled with possibly fat-tailed distribution for the returns, results into superior VaR prediction. Li and Lin,[65] for example, estimated a Markov switching ARCH model and showed that it outperforms both ARCH and GARCH in in- and out-sample VaR predictions. Chen et al[57] have considered a double Markov switching GARCH model that is, in spirit, similar to the model proposed here and showed that their proposed model is superior to standard approaches in VaR estimation. Yu et al,[66] for example, modeled the VaR measure indirectly by using

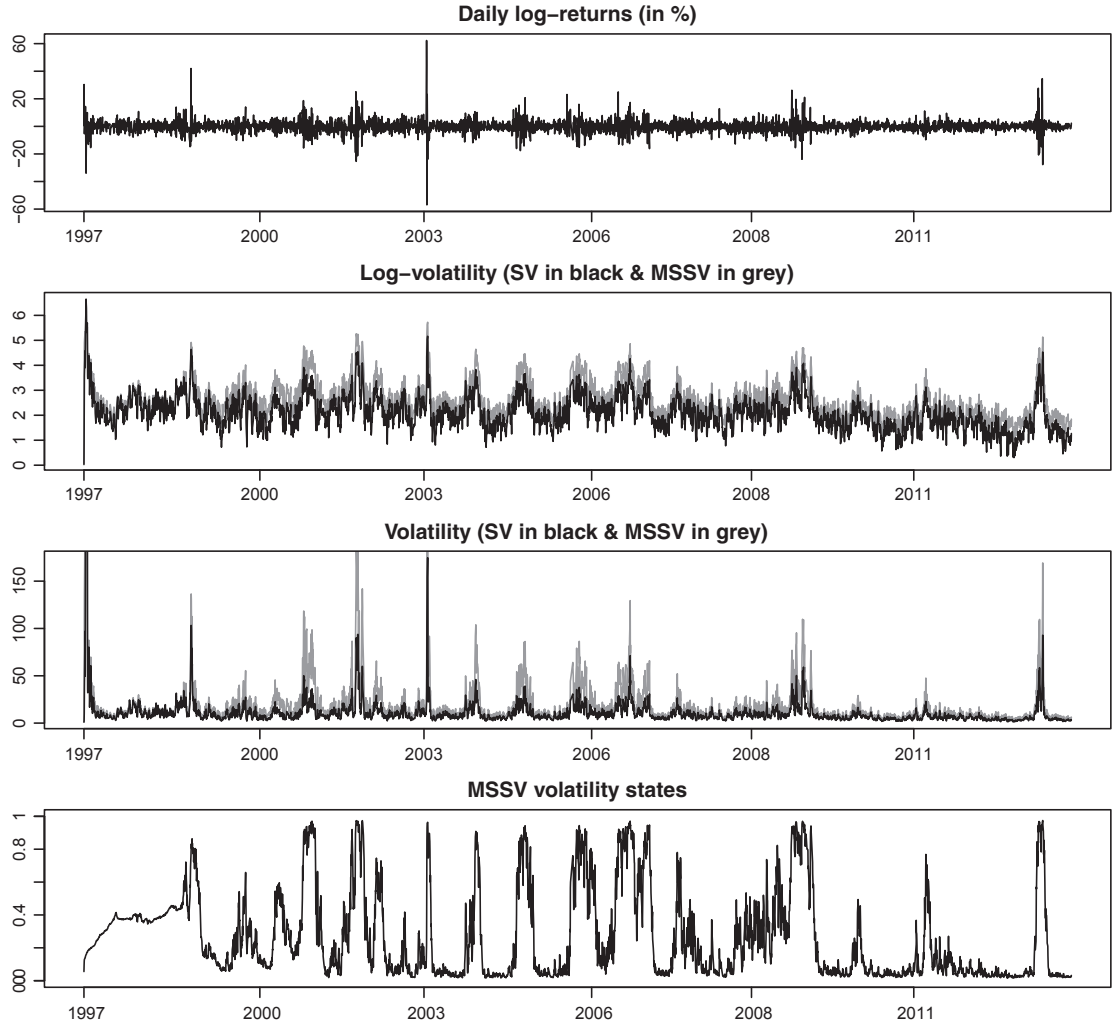

FIGURE 8 Filtered volatilities and volatility states (the mean probability of being in a state one $s_t = 1$) for natural gas data for SV-DPM and MSSV-DPM models. DPM, Dirichlet process mixture; MSSV, Markov switching stochastic volatility; SV, stochastic volatility

threshold (nonlinear) GARCH models and found that the performance of their proposed approach outperforms other methods.

5 | DISCUSSION

This paper augments the existing SV-DPM model with Markov switching jumps to capture different volatility regimes, resulting into an MSSV-DPM model. We test the newly proposed model on simulated data and find that the PL estimation procedure is able to identify different volatility regimes. We present a real data application using three financial time series of the returns for one index (S&P500), one company (Ford), and one commodity (natural gas). We find that the MSSV-DPM model performs significantly better than the SV-DPM model if we consider the entire predictive distribution of the returns for Ford and natural gas data, but not for the S&P500 data. If we consider the tails of the distributions, the MSSV-DPM model significantly outperforms the SV-DPM model for the three data sets for all percentiles (1, 5, and 10) of the tail. Finally, the volatility persistence parameter estimates drop significantly after including the Markov switching specification, a finding in line with the results in multiple previous studies. Overestimation of the volatility persistence leads to incorrect understanding of the predictability of the volatility.

A possible extension to the proposed model could be including a regime switch not only in the intercept term but also in the persistence parameter, and/or in the mean equation for the returns (which is now assumed to be zero), same as in the work of Chen et al.[57] This would allow for even more flexibility in the nonlinear volatility dependence process.

Another future research line could deal with comparing the model proposed in the manuscript to a semiparametric regime switching GARCH model (in other words, its GARCH counterpart), adding a contribution to a large class of literature dealing with the on-going debate of SV vs. GARCH.

REFERENCES

1. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. 1982;50(4):987-1008.
2. Taylor SJ. Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices. In: Anderson OD, ed. *Time Series Analysis: Theory and Practice 1*. Amsterdam, The Netherlands: North-Holland; 1982:203-226.
3. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *J Econ*. 1986;31(3):307-327.
4. Broto C, Ruiz E. Estimation methods for stochastic volatility models: a survey. *J Econ Surv*. 2004;18(5):613-650.
5. Jacquier E, Polson NG, Rossi PE. Bayesian analysis of stochastic volatility models. *J Bus Econ Stat*. 1994;12(4):371-389.
6. Kim S, Shephard N, Chib S. Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev Econ Stud*. 1998;65(3):361-393.
7. Taylor SJ. *Modelling Financial Time Series*. 1st ed. New York, NY: John Wiley & Sons; 1986.
8. Taylor SJ. Modeling stochastic volatility: a review and comparative study. *Math Financ*. 1994;4(2):183-204.
9. Abanto-Valle CA, Bandyopadhyay D, Lachos VH, Enriquez I. Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Comput Stat Data Anal*. 2010;54(12):2883-2898.
10. Chib S, Nardari F, Shephard N. Markov chain Monte Carlo methods for stochastic volatility models. *J Econ*. 2002;108(2):281-316.
11. Jacquier E, Polson NG, Rossi PE. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J Econ*. 2004;122(1):185-212.
12. Gallant A, Hsieh D, Tauchen G. Estimation of stochastic volatility models with diagnostics. *J Econ*. 1997;81(1):159-192.
13. Harvey AC, Ruiz E, Shephard N. Multivariate stochastic variance models. *Rev Econ Stud*. 1994;61(2):247-264.
14. Nakajima J, Omori Y. Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Comput Stat Data Anal*. 2009;53(6):2335-2353.
15. Sandmann G, Koopman SJ. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *J Econ*. 1998;87(2):271-301.
16. Barndorff-Nielsen OE. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand J Stat*. 1997;24(1):1-13.
17. Mahieu RJ, Schotman PC. An empirical application of stochastic volatility models. *J Appl Econ*. 1998;13(4):333-360.
18. Liesenfeld R, Richard J. Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. *Econ Rev*. 2006;25(2-3):335-360.
19. Jensen MJ, Maheu JM. Bayesian semiparametric stochastic volatility modeling. *J Econ*. 2010;157(2):306-316.
20. Jensen MJ, Maheu JM. Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *J Econ*. 2014;178(Pt 3):523-538.
21. Delatola E-I, Griffin JE. Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Anal*. 2011;6(4):901-926.
22. Delatola E-I, Griffin JE. A Bayesian semiparametric model for volatility with a leverage effect. *Comput Stat Data Anal*. 2013;60:97-110.
23. Zaharieva MD, Trede M, Wilfling B. *Bayesian Semiparametric Multivariate Stochastic Volatility With an Application to International Volatility Co-Movements*. Müenster, Germany. Center for Quantitative Economics, University of Müenster; 2017. Working Papers.
24. Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann Stat*. 1974;2(6):1152-1174.
25. Ferguson TS. Bayesian density estimation by mixtures of normal distributions. In: Rizvi MH, Rustagi J, Seigmund D, eds. *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*. Vol. 24. New York, NY: Academic Press; 1983:287-302.
26. Lo AY. On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann Stat*. 1984;12(1):351-357.
27. Tokdar ST. Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhyā Indian J Stat*. 2006;8(Pt 41):90-110.
28. Mencía J, Sentana E. Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *J Econ*. 2009;153(2):105-121.
29. Titterington D, Smith A, Makov U. *Statistical Analysis of Finite Mixture Distributions*. Chichester, UK: Wiley; 1985.
30. Lau JW, So MKP. Bayesian mixture of autoregressive models. *Comput Stat Data Anal*. 2008;53(1):38-60.
31. Lau JW, So MK. A Monte Carlo Markov chain algorithm for a class of mixture time series models. *Stat Comput*. 2011;21(1):69-81.
32. Vo MT. Regime-switching stochastic volatility: evidence from the crude oil market. *Energy Econ*. 2009;31(5):779-788.
33. So MKP, Lam K, Li WK. A stochastic volatility model with Markov switching. *J Bus Econ Stat*. 1998;16(2):244-253.
34. Kalimipalli M, Susmel R. Regime-switching stochastic volatility and short-term interest rates. *J Empir Finance*. 2004;11(3):309-329.
35. Shibata M, Watanabe T. Bayesian analysis of a Markov switching stochastic volatility model. *J Jpn Stat Soc*. 2005;35(2):205-219.
36. Carvalho CM, Lopes HF. Simulation-based sequential analysis of Markov switching stochastic volatility models. *Comput Stat Data Anal*. 2007;51(9):4526-4542.

37. Chen CWS, So MKP, Liu F-C. A review of threshold time series models in finance. *Stat Its Interface*. 2011;4(2):167-181.
38. Franses PH, van Dijk D. *Non-Linear Time Series Models in Empirical Finance*. Cambridge, UK: Cambridge University Press; 2000.
39. Li WK. The threshold approach in volatility modelling. In: *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*. Singapore: World Scientific Publishing Co Pte Ltd; 2009:95-100.
40. Virbickaite A, Ausín MC, Galeano P. Bayesian inference methods for univariate and multivariate Garch models: a survey. *J Econ Surv*. 2015;29(1):76-96.
41. Lopes HF, Polson NG. Extracting S&p500 and NASDAQ Volatility: the credit crisis of 2007-2008. In: O'Hagan T, West M, eds. *The Oxford Handbook of Applied Bayesian Analysis*. New York, NY: Oxford University Press; 2010:319-342.
42. Storvik G. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Trans Signal Process*. 2002;50(2):281-289.
43. Carvalho CM, Johannes MS, Lopes HF, Polson NG. Particle learning and smoothing. *Stat Sci*. 2010;25(1):88-106.
44. Warty SP, Lopes HF, Polson NG. Sequential Bayesian learning for stochastic volatility with variance-gamma jumps in returns. *Appl Stoch Model Bus Ind*. 2018;34(4):460-479.
45. Virbickaitė A, Lopes HF, Ausín MC, Galeano P. Particle learning for Bayesian semi-parametric stochastic volatility model. *Econ Rev*. 2019. https://doi.org/10.1080/07474938.2018.1514022.
46. Omori Y, Chib S, Shephard N, Nakajima J. Stochastic volatility with leverage: fast and efficient likelihood inference. *J Econ*. 2007;140(2):425-449.
47. Carter CK, Kohn R. On Gibbs sampling for state space models. *Biometrika*. 1994;81(3):541-553.
48. Fruhwirth-Schnatter S. Data augmentation and dynamic linear models. *J Time Ser Anal*. 1994;15(2):183-202.
49. Ausín MC, Galeano P, Ghosh P. A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *Eur J Oper Res*. 2014;232(2):350-358.
50. Jensen MJ, Maheu JM. Bayesian semiparametric multivariate GARCH modeling. *J Econ*. 2013;176(1):3-17.
51. Kalli M, Walker SG, Damien P. Modeling the conditional distribution of daily stock index returns: an alternative Bayesian semiparametric model. *J Bus Econ Stat*. 2013;31(4):371-383.
52. Virbickaitė A, Ausín MC, Galeano P. A Bayesian non-parametric approach to asymmetric dynamic conditional correlation model with application to portfolio selection. *Comput Stat Data Anal*. 2016;100:814-829.
53. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *J Am Stat Assoc*. 1995;90(430):577-588.
54. Lopes HF, Carvalho CM. Factor stochastic volatility with time varying loadings and Markov switching regimes. *J Stat Plan Inference*. 2007;137(10):3082-3091.
55. Liu J, West M. Combined parameter and state estimation in simulation-based filtering. In: Doucet A, de Freitas N, Gordon N, eds. *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer-Verlag New York; 2001:197-223.
56. Abanto-Valle CA, Migon HS, Lopes HF. Bayesian modeling of financial returns: a relationship between volatility and trading volume. *Appl Stoch Model Bus Ind*. 2010;26(2):172-193.
57. Chen CWS, So MKP, Lin EMH. Volatility forecasting with double Markov switching GARCH models. *J Forecast*. 2009;28(8):681-697.
58. Lopes HF, Carvalho CM, Johannes MS, Polson NG. Particle learning for sequential Bayesian computation. In: Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AF, West M, eds. *Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting*. New York, NY: Oxford University Press; 2011:317-360.
59. Lo AY. Weighted Chinese restaurant process. *COSMOS*. 2005;1(1):107-111.
60. Carvalho CM, Lopes HF, Polson NG, Taddy MA. Particle learning for general mixtures. *Bayesian Anal*. 2010;5(4):709-740.
61. Chopin N, Iacobucci A, Marin J-M, et al. On particle learning: comments on "Particle learning for sequential Bayesian computation" by Lopes, Carvalho, Johannes, and Polson. In: Bernardo JM, Bayarri MJ, Berger JO, et al, eds. *Bayesian Statistics 9*. Oxford University Press, UK: Oxford University Press; 2011:317-360. 2011.
62. Frühwirth-Schnatter S. *Finite Mixture and Markov Switching Models*. New York, NY: Springer Science+Business Media LLC; 2006. *Springer Series in Statistics*.
63. Koop G. *Bayesian Econometrics*. Chichester, UK: John Wiley & Sons; 2003.
64. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90(430):773-795.
65. Li M-YL, Lin H-wW. Estimating value-at-risk via Markov switching ARCH models - an empirical study on stock index returns. *Appl Econ Lett*. 2004;11(11):679-691.
66. Yu PLH, Li WK, Jin S. On some models for value-at-risk. *Econ Rev*. 2010;29(5-6):622-641.

How to cite this article: Virbickaitė A, Lopes HF. Bayesian semiparametric Markov switching stochastic volatility model. *Appl Stochastic Models Bus Ind*. 2019;35:978–997. https://doi.org/10.1002/asmb.2434

APPENDIX

PL FOR MSSV-DPM

1. **Resampling.** Resample old particles (parameters and the set of sufficient statistics, including the three state variables, h_t, k_t, and λ_t) with weights proportional to the predictive density of the returns, which can be obtained as follows, where $p(r_t|h_t, \Theta)$ and $p(h_t|h_{t-1}, \Theta)$ are as in (6) and (7):

$$\begin{aligned} p(r_t|h_{t-1}, \Theta) &= \int p(r_t|h_t, \Theta)p(h_t|h_{t-1}, \Theta)dh_t \\ &= \frac{1}{c+t-1}\sum_{j=0}^{L^\star_{t-1}} n_j \int f_{\mathcal{N}}\left(r_t; h_t + \mu_j, \sigma_j^2\right) f_{\mathcal{N}}\left(h_t; \gamma_0 + \gamma_1\lambda_t + \beta h_{t-1}, \tau^2\right) dh_t \\ &= \ldots \int \frac{\exp\left\{-(r_t - (h_t + \mu_j))^2/(2\sigma_j^2)\right\}}{\sqrt{2\pi\sigma_j^2}} \frac{\exp\left\{-(h_t - (\gamma_0 + \gamma_1\lambda_t + \beta h_{t-1}))^2/(2\tau^2)\right\}}{\sqrt{2\pi\tau^2}} dh_t \\ &= \frac{1}{c+t-1}\sum_{j=0}^{L^\star_{t-1}} n_j f_N\left(r_t; \gamma_0 + \gamma_1\lambda_t + \beta h_{t-1} + \mu_j, \tau^2 + \sigma_j^2\right), \end{aligned}$$

   where $n_0 = c$ and $(\gamma_0, \gamma_1, \beta, \tau^2, p, q, \mu_1, \ldots, \mu_{L^\star_{t-1}}, \sigma_1^2, \ldots, \sigma^2_{L^\star_{t-1}})$ have been simulated at the end of the previous period.
2. **Sampling.** In this step, we propagate the latent states h_t, the latent volatility states λ_t, and the indicator variables k_t, which indicate to which mixture component the observation belongs to. Note that the tilde above the parameter indicates that the particle has been resampled in the first step.
   (a) The volatility state variable λ_t is propagated according to the following:

$$\begin{aligned} p\left(\lambda_t|\tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}, r_t\right) &\propto p\left(r_t|\tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}\right) p\left(\lambda_t|\tilde{\lambda}_{t-1}\right) \\ \lambda_t|\tilde{\lambda}_{t-1}, \tilde{h}_{t-1}, \tilde{\Theta}, r_t &\sim \mathcal{BER}\left(\frac{z_2}{z_1 + z_2}\right), \end{aligned}$$

   where z_1 and z_2 as in (9) and $\tilde{\alpha} = \tilde{\gamma}_0 + \tilde{\gamma}_1\lambda_t$.
   (b) For sampling, h_t make use of $p(h_t|\tilde{h}_{t-1}, r_t, \lambda_t, \tilde{\Theta}) \propto p(r_t|h_t, \tilde{\Theta})p(h_t|\tilde{h}_{t-1}, \lambda_t, \tilde{\Theta})$, where $p(r_t|h_t, \Theta)$ and $p(h_t|h_{t-1}, \lambda_t, \Theta)$ are as in (6) and (7)

$$\begin{aligned} p\left(h_t|\tilde{h}_{t-1}, \tilde{\Theta}, \tilde{n}, \tilde{L}^\star_{t-1}, \lambda_t, r_t\right) &\propto \sum_{j=0}^{\tilde{L}^\star_{t-1}} \frac{\tilde{n}_j}{c+t-1} f_{\mathcal{N}}\left(r_t; h_t + \tilde{\mu}_j, \tilde{\sigma}_j^2\right) f_{\mathcal{N}}\left(h_t; \tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1}, \tilde{\tau}^2\right) \\ h_t|\tilde{h}_{t-1}, \tilde{\Theta}, \tilde{n}, \tilde{L}^\star_{t-1}, \lambda_t, r_t &\sim \sum_{j=0}^{L^\star_{t-1}} \frac{\tilde{n}_j}{c+t-1} \mathcal{N}(h_t; m_{hj}, V_{hj}), \end{aligned}$$

   where $V_{hj} = A_j\tilde{\sigma}_j^2$, $m_{hj} = A_j(r_t - \tilde{\mu}_j) + (1 - A_j)(\tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1})$, $A_j = \tilde{\tau}^2/(\tilde{\tau}^2 + \tilde{\sigma}_j^2)$, and $\tilde{\alpha} = \tilde{\gamma}_0 + \tilde{\gamma}_1\lambda_t$.
   (c) For sampling new indicators k_t, make use of $p(k_t = j|r_t, \tilde{h}_{t-1}, \tilde{\Theta}) \propto p(r_t|k_t = j, \tilde{h}_{t-1}, \tilde{\Theta})p(k_t = j|\tilde{h}_{t-1}, \tilde{\Theta})$, where

$$p\left(r_t|k_t = j, \tilde{h}_{t-1}, \tilde{\Theta}\right) = \int p\left(r_t|h_t, \tilde{\Theta}\right) p\left(h_t|\tilde{h}_{t-1}, \tilde{\Theta}\right) dh_t$$

   and

$$p\left(k_t = j|\tilde{h}_{t-1}, \tilde{\Theta}\right) \propto \frac{\tilde{n}_j}{c+t-1},$$

   therefore,

$$p\left(k_t = j|r_t, \tilde{h}_{t-1}, \tilde{\Theta}\right) \propto \tilde{n}_j f_N\left(r_t; \tilde{\alpha} + \tilde{\beta}\tilde{h}_{t-1} + \tilde{\mu}_j, \tilde{\tau}^2 + \tilde{\sigma}_j^2\right), \; j = 1, \ldots, L^\star_{t-1} + 1,$$

   where $\tilde{n}_{L^\star_{t-1}+1} = c$ and $\sigma^2_{L^\star_{t-1}+1} = \sigma_0^2$.

3. **Propagating sufficient statistics and learning Θ.**
(c.3) and (c.4) sampling τ^2 and β, ie,

$$
\begin{aligned}
p(\beta, \tau^2|h_t) &\propto p(h_t|\beta, \tau^2)p(\beta, \tau^2) \\
&\propto f_{\mathcal{N}}\left(h_t; \tilde{\alpha} + \beta\tilde{h}_{t-1}, \tau^2\right) f_{\mathcal{TN}_{(-1,1)}}\left(\beta; \tilde{m}_\beta, \tilde{V}_\beta\tau^2\right) f_{\mathcal{IG}}\left(\tau^2; \tilde{b}_0/2, \tilde{b}_0\tilde{\tau}_0^2/2\right) \\
\tau^2 &\sim \mathcal{IG}\left(\tau^2; \frac{\tilde{b}_0+1}{2}, \frac{\tilde{b}_0\tilde{\tau}_0^2 + \left(\tilde{m}_\beta\tilde{h}_{t-1} - (h_t - \tilde{\alpha})^2\right) / \left(1 + \tilde{V}_\beta\tilde{h}_{t-1}^2\right)}{2}\right) \\
\beta &\sim \mathcal{TN}_{(-1,1)}\left(\beta; \frac{\tilde{m}_\beta + \tilde{V}_\beta\tilde{h}_{t-1}(h_t - \tilde{\alpha})}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2}, \frac{\tilde{V}_\beta\tau^2}{1 + \tilde{V}_\beta\tilde{h}_{t-1}^2}\right).
\end{aligned}
$$

Sufficient statistics updates and sampling for the rest of the parameters are analogous to (c.3) and (c.4).